ARI Research Note 88-38

AD-A195 621

DTIC FILE COPY

USE OF A WORKING MODEL IN FAULT DIAGNOSIS

Kenneth R. Allison
Georgia Institute of Technology

for

Contracting Officer's Representative
Judith Orasanu

BASIC RESEARCH LABORATORY
Michael Kaplan, Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

January 1988

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Technical review by

Dan Ragland

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC COPY INSPECTED 4

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARI Research Note 88-38	2. GOVT ACCESSION NO. - -	3. RECIPIENT'S CATALOG NUMBER - -
4. TITLE (*and Subtitle*) Use of a Working Model in Fault Diagnosis		5. TYPE OF REPORT & PERIOD COVERED Interim Report July 86 - June 87
		6. PERFORMING ORG. REPORT NUMBER - -
7. AUTHOR(s) Kenneth R. Allison		8. CONTRACT OR GRANT NUMBER(s) MDA903-86-C-0173
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Information and Computer Science Georgia Institute of Technology Atlanta, GA 30332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue Alexandria, VA 22333-5600		12. REPORT DATE January 1988
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS (*if different from Controlling Office*) - -		15. SECURITY CLASS. (*of this report*) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE n/a

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

- -

18. SUPPLEMENTARY NOTES

Judith Orasanu, contracting officer's representative

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

Cognitive Science
Decision Making
Problem Solving
Mental Models
Judgement

(This document [illegible])

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

Many types of knowledge are required for effective reasoning in the diagnostic domain. In particular, knowledge about the normal functioning of a system is essential in order to troubleshoot the system. ~~In this research note, we define~~ a working model which represents a troubleshooter's integrated knowledge of system components, to include input, output, structure, function, and causal relationships.

Two ways that the working model can aid in fault diagnosis are*(OVER)

ARI RESEARCH NOTE 88-38 This paper

20. Abstract (continued)

in generating hypotheses for subsequent testing and in verifying or explaining faulty behavior. ~~In this paper, we present~~ a representation of an automobile mechanic for the mental working model. The emphasis in this domain is to use the working model to generate new hypotheses, in a manner consistent with the behavior of real mechanics.

Keywords: Cognitive science; problem solving; decision making; mental models; judgement; Army research (KR).

Use Of A Working Model In Fault Diagnosis*

Kenneth R. Allison

School Of Information And Computer Science
Artificial Intelligence Research Group
Georgia Institute Of Technology
Atlanta, GA 30332
(404) 894-5550
E-mail: kra@ics.gatech.edu

ABSTRACT

Effective reasoning in a diagnostic domain requires many types of knowledge. In particular, knowledge about the normal functioning of a system is crucial to the ability to troubleshoot the system. We define a working model that represents a troubleshooter's integrated knowledge about system components, including input, output, structure, function, and causal relationships. Two ways the working model can aid fault diagnosis are
(1) in generating hypotheses for subsequent testing, and
(2) in verifying or explaining faulty behavior.
In this paper, we present a representation for the mental working model of an automobile mechanic. Our emphasis in this domain is to use the working model to generate new hypotheses, in a manner consistent with the behavior of real mechanics.

BACKGROUND AND MOTIVATION

One of our current research projects here at Georgia Tech is an investigation into the reasoning and problem solving processes used by an automobile mechanic while trying to repair a car. This is part of a more general attempt to discover the differences in problem solving techniques between novices and experts. During our study, we have taken live protocols of the problem-solving behavior of students in auto repair at a local vocational-technical institute. The students were divided into four categories: novice, intermediate, advanced, and expert (the instructor). The protocols were taken while the students were trying to diagnose cars into which we had previously introduced a fault. We coded the protocols, looking for evidence about how hypotheses are generated. This paper represents a first attempt at codifying our theories about how new hypotheses are generated; we plan on building a computer program that simulates the diagnostic behavior of an automobile mechanic.

AVAILABLE KNOWLEDGE

During our review of the protocols, we noticed that knowledge used in generating hypotheses seemed to come from three main sources:

* This research is supported in part by the Army Research Institute under Contract No. MDA-903-86-C-173 and in part by the Georgia Institute of Technology.

(1) a set of symptom-fault pairs
(2) a working model
(3) manuals (or other external sources).

A symptom-fault pair is a simple association between an observable symptom and a potential fault that could cause that symptom to be manifested. An example would be: If the car won't start (*symptom*), the battery may be dead (fault). There may be multiple faults associated with a particular symptom. If this is the case, there is also probability information about how likely each one is of being the actual culprit. This probability information, presumably compiled over many cases, is what allows the mechanic to check for the more common fault condition first. This often leads to very rapid diagnosis on typical cases.

The working model is basically the mechanic's mental model of a normally functioning car. This includes information about specific components and about (sub)systems. An example component would be the battery, and an example system would be the ignition system. The working model also contains knowledge about how the components (and systems) are interconnected. The working model is hierarchical: components, for instance, can be composed of subcomponents; these subcomponents can also be viewed as top level components at a lower level.

Knowledge used in diagnosis also came from external sources, such as the diagnostic "trouble tree" found in some repair manuals, or advice from a more experienced mechanic (a hint from the instructor, in our protocols). This information is usually needed only when the mechanic is stumped, or realizes that he doesn't have the necessary knowledge to deal with a certain component or test procedure. However, if the knowledge is complex, and readily available in a book (such as the testing procedure for the Electronic Control Module (ECM) on newer cars), the mechanic may make no effort to memorize it.

These three knowledge sources were used in the generation of hypotheses during actual problem solving episodes; however, other knowledge was used in the formation of the symptom-fault pairs and the working model. The students we studied were taking classes in auto repair, so much of the initial working model, and some of the symptom-fault pairs, came from classroom learning. Experience plays an important role as well; mechanics who have worked on hundreds of cars have refined their knowledge (working model and especially symptom-fault) as a result. "Some things you can't learn from books" - a well worn saying, but true.

THE WORKING MODEL: REPRESENTATION

The working model is represented in our program using frames, which allow easy inheritance in hierarchical representations. This is important because the working model divides the knowledge of the topology of the car into both functional and structural hierarchies. Thus, the structural hierarchy allows an individual electrical wire to be an instance of a more general electrical wire frame, or the fuel pump frame to inherit properties of a prototypical pump. The functional hierarchy, on the other hand, divides the car into systems, components, and subcomponents.

A system is a series of interconnnected components that achieves a higher-level goal. The system components are instrumental to the achievement of the system goal. An example system is the ignition system, which has the following components: ignition-switch, starter, and battery. These parts are connected by electrical wires, which is a component in its own right (conduit).

A component is an average, everyday part which one could walk into the auto parts store and buy. It is a separate, replacable part that can be lifted as one piece. A component may have subcomponents inside it or otherwise attached to it as integral parts.

The difference between components and subcomponents is a grey area at times. A "component" is part of a system; a "subcomponent" is instrumental to the functioning of a component. A subcomponent is an integral part of a component. For instance, the fuel pump motor is a subcomponent of the fuel pump, but the fuel tank is not. Another heuristic for deciding borderline cases is whether the faulty part is replaced as a unit. Generally, fuel pumps are replaced as a whole, instead of taking them apart to replace a faulty subcomponent such as the fuel pump motor.

As an example, here is an English representation of the fuel pump:

FUEL-PUMP

Subcomponents: fuel-pump-sensor, fuel-pump-motor
Part-of: fuel-system
Input: fuel FROM fuel-tank VIA fuel-line
Output: fuel TO carburetor/fuel-injection-unit VIA fuel-line
Connected-to: fuel-tank VIA fuel-line
carburetor/fuel-injection-unit VIA fuel-line
Test: Sound FOR 2-3 seconds WHEN key is turned
Function: Move fuel from fuel-tank to engine against gravity

THE WORKING MODEL: FUNCTIONS

Mechanics at all levels of expertise appear to use symptom-fault pairs to generate initial hypotheses. The initial symptom is generally the customer's complaint (reason for bringing the car in for repair). For example, the initial symptom might be that the car won't start, or that it stalls frequently. The mechanic will usually try to verify the complaint first, in case the customer is mistaken about the symptom or has omitted another symptom. After this step, the mechanic has an initial symptom set available as a starting point for diagnosis. The symptom-fault knowledge set is probed with the initial symptom, and the resultant set of potential faults becomes the initial hypothesis set. One of the hypotheses is chosen (by probability of failure and ease of testing) as the current hypothesis. This current hypothesis either points to a bad system (e.g., problem is in the fuel system), or a bad component (e.g., battery is run down). Diagnostic reasoning then proceeds at either the system level or the component level. By diagnostic reasoning, we simply mean the problem-solving and reasoning strategies used by mechanics to diagnose the fault (identify the faulty component). The following paragraphs explain the diagnostic reasoning at the system and component levels.

1. System-level reasoning

If the mechanic is pointed to a faulty system, the next step is to isolate the faulty component within the system. This means that one of the system components should become the next hypothesis. The first attempt to choose the component to focus on next is made by again trying the symptom-fault knowledge base, this time using the faulty system as the symptom. This may yield the desired component-level hypothesis. For example, the symptom "bad fuel system" may have "worn out fuel pump" as its associated fault. The fuel pump then becomes the new hypothesis.

Another way to choose the component to focus on next is to start at the endpoint of a system. In the absence of specific symptom-fault knowledge, the system endpoint is a suitable default. In most cases, unless the mechanic is a rank novice, the symptom-fault knowledge provides a suitable hypothesis. As a mechanic gains experience by working on many different cases, he forms new symptom-fault associations. Thus, an expert mechanic who has seen thousands of cases has a very complete and highly accurate set of symptom-fault associations. Defaulting to the system endpoint to get a component-level hypothesis is therefore only applicable to a beginner.

At this point in diagnosis, a system is considered faulty, and a candidate component within the system is the current hypothesis. Starting with this "focus" (component hypothesis), a trace within the system can be done until the faulty component is discovered. The system trace starts with the examination of the outputs of the focus. As explained elsewhere in this paper, a component is only confirmed as being faulty when it gives incorrect output while receiving all correct inputs. Incidentally, this is one of the differences between novices and expert. A novice is content to confirm a hypothesis if the output is incorrect, and not even bother with inputs. In one of the protocols, a novice switches on the key to listen for the fuel pump to run. Because the fuel pump makes no sound (incorrect output), the novice confidently proclaims that the fuel pump is broken, and would presumably have replaced it if this was a real case. However, the real problem was that the fuel pump fuse was burned out. This meant that electrical power was not reaching the fuel pump motor (incorrect input). The more advanced students solved this case correctly because of their superior diagnostic strategies at the system level. The novice algorithm is:

1. Check outputs of the component in question (current focus).

2. If all outputs are correct, all system components leading up to the current focus are OK. RETURN.

3. If an output is incorrect, the component is faulty. RETURN.

The expert algorithm is:

1. Check outputs of the component in question (current focus).

2. If all outputs are correct, all system components leading up to the current focus are OK. RETURN.

3. If an output is incorrect, check inputs to the component.

4. If all inputs are correct, the component is faulty (see component level reasoning for an exception). RETURN.

5. If all inputs are not correct, use the working model to trace back in the system to the component responsible for that input. This component becomes the new focus. REPEAT ALGORITHM.

Although the expert algorithm seems simplistic, it is important to note that novices do not always understand the reasoning behind it. This knowledge is crucial to a correct diagnosis in many cases. Another note is that experts rarely have to do a long system trace because of their extensive symptom-fault set. However, experts can do these traces, and do if they are trying to diagnose a fault in some unfamiliar part of the system.

2. Component-level reasoning

Reasoning can also occur within a component, because some components have separate subcomponents as integral parts, as explained earlier. For example, the fuel pump contains the fuel pump motor as a subcomponent. The distinction between the two is admittedly fuzzy at times, but the motivation behind it is that a mechanic will usually stop at a certain point in diagnosis, and replace the faulty component. It is generally more cost-effective to replace the battery, for instance, even though it is probably a single dead cell causing the problem. Once a component is verified as being faulty, it is replaced; repair or replacement of the subcomponent actually causing the problem is not attempted.

What good are subcomponents then? A mechanic still has knowledge about them, and can use them in reasoning about the components. This can sometimes lead to new hypotheses. To elaborate on the earlier example, the advanced mechanics knew about the fuel pump motor subcomponent of the fuel pump. They knew that the fuel pump motor is what normally makes the noise when the key is turned. They also knew that the fuel pump motor requires electrical energy to run. This led them to the actual fault, the fuel pump fuse.

THE WORKING MODEL: ASSUMPTIONS

As in any system, there are certain underlying assumptions that are necessary in order to be able to make valid inferences. Some of the assumptions for the working model follow.

1. All components in a system must be working properly for the whole system to work. Thus, if a system has components and flow 1 -> 2 -> 3, then a precondition for 1 to work properly is also a precondition for 2 and 3. Take the fuel system as an example. The flow of fuel is fuel tank -> fuel pump -> ... -> cylinders. A precondition for every component is that there is fuel in the fuel tank, or else the component is "not working" in some sense. However, we don't want to diagnose every component in the fuel system as faulty if the fuel tank is empty. In using the working model for diagnosis, a component is tested by seeing if it produces normal outputs when given normal inputs. Obviously, the inputs will not be "normal" unless the components and connections in the system leading up to it are all working properly. However, a normal input can be sometimes be fed into a component directly, bypassing any faulty connections, and thus allowing a component to be tested. Therefore, the only preconditions for a component are that it receives the proper inputs.

2. The normal condition for a component is that it is clean, not corroded or cracked, and has no missing subcomponents (parts). A clog in a fuel line is a fault that affects the output of the fuel line, but the knowledge that there must be no clogs for a properly functioning fuel line seems to belong more in the symptom-fault knowledge base than in the normal working model.

3. For the car to run perfectly, all of the car's systems must be working properly. Some systems are of higher criticality than others (brakes vs air conditioning), but this type of knowledge doesn't appear to be too important in a diagnostic domain. If a person brings a car in complaining about the air conditioner being broken, the mechanic will try to fix the air conditioner. The brakes will not be checked.

4. It is the purpose of a (sub)system to achieve a (sub)goal necessary to the proper working of the car. The goal of a system is referred to as a function in this model. To achieve its goal (perform its function), a system must almost always transport some substance (matter or energy) from one point to another. It is assumed that no change is made to the substance except physical location unless noted as a "Result" of a function. Another assumption: Given components 1 -> 2 -> 3, with the arrows indicating the flow of a substance, less substance is available to 1, and more substance is available to 3, as a result of the substance passing through 2.

RELATED WORK

Early AI work in troubleshooting was mostly in the medical diagnosis domain (e.g. MYCIN (Shortliffe 1976)) and relied very heavily on symptom-fault sets. Because many parts of the body are inaccessible, this approach is often appropriate in medical reasoning, and appears to be how doctors arrive at a diagnosis. Kuipers (1986) has pointed out that doctors do use some causal reasoning, and has been working on modeling physiological mechanisms, but he notes that the hypothesis-driven (symptom-fault) approach predominates in real physicians. Mechanics, however, appear to symptom-fault information and causal reasoning on a working model in roughly equal proportions. This is because, in the automobile repair domain, the working model is fully explainable, and the car components are accessible for testing. Causal reasoning (using a working model) in this domain is both applicable and beneficial.

Representation of physical objects in a principled way that allows straightforward reasoning has been a strong area of research over the last few years. Most papers propose various hierarchies of objects, and the objects are usually portrayed in a frame-like manner. Our representations use an eclectic mix of ideas from the work of de Kleer & Brown (1981), Forbus & Gentner (1986), Kuipers (1984), and Lehnert (1978) for the working model's representation.

Causal reasoning, especially in the area of qualitative physics, has been a strong research area recently. Bobrow (1984) gives a good overview of this field. Causal reasoning about a working model is necessary in the mechanics domain, but not to the the depth proposed by de Kleer and Brown (1981). This work is too detailed for our present needs. Our level of detail is mostly at the component level, and is concerned mainly with which connections exist among components, and the flow of substances between them. The detailed structural representation of every component is not needed. To represent complex ideas such as the combustion cycle, the aggregation techniques of Weld (1986) are probably more appropriate for our purposes. It appears that the main purpose of diagnosis in the automobile repair domain is to find the faulty part and then replace it. The added capability of being able to use envisionment (de Kleer & Brown 1981) to explain in detail how the faulty part caused the external behavior is not really necessary in a diagnostic domain (Sembugamoorthy & Chandrasekaran 1986, p. 67).

Hunt (1981) wrote a rule-based diagnosis program called FAULT that modeled the user's knowledge in a standard production system. Although our program is not rule-based, some of the same knowledge is needed in both pprograms because of the nature of the domain. Thus, our symptom-fault pairs correspond to Hunt's S-rules (symptomatic search rules), and some of our assumptions about when a part is faulty correspond to his T-rules (topographic search rules).

DIRECTIONS FOR FUTURE RESEARCH

Implementing our theories fully on a computer is the next step. The program is an attempt to simulate the problem-solving behavior of a mechanic. The mechanic protocols are a valuable source of feedback on the accuracy of our program's behavior. At the same time, computer simulation forces our theories to be well-defined.

Another set of protocols is currently being taken. This time, a knowledge assessment is being done on the subjects before and after the new series of protocols. Because some of the problems will be similar, we expect to see the effects of learning on the solution to the second encounter with the problem. How the knowledge and experience gained on the first attempt are incorporated will be valuable clues to the underlying learning and reasoning mechanisms. In addition, there will be a debriefing session after each protocol to review the just completed problem-solving session. This debriefing will allow a more in-depth examination of the reasoning processes used in the problem solution, while not distracting the subject from the ongoing task. We hope to bring out some of the knowledge and problem-solving strategies that are being used, but not verbalized in the protocol. For instance, we expect that more case-based reasoning is being used than is apparent from the protocols alone.

The computer implementation is in an early stage of development. However, progress is being made, and the program will become more sophisticated as work continues. An eventual goal is to have the program become part of a tutoring system for novice mechanics. Two aspects of our approach give the potential for a very sophisticated system. First, by making our program's knowledge representation and problem solving strategies match a person's, the system will more easily be able to explain its behavior. Similarly, it will be easier to model the student's knowledge, which is crucial to good tutoring (e.g. Burton, 1982; Sleeman, 1982).

Secondly, if the program stores some type of history of its previous failures, it can recognize similar mistakes on the part of the student, and be able to deal with them effectively. One way this could be done is in a case-based reasoning system (Kolodner et al., 1985; Simpson, 1985) in which a program with an evolving, dynamic memory remembers previous failures in addition to the way they were eventually resolved.

Many open research questions remain. Getting the program to learn (i.e. evolve from novice to expert) will be very difficult. For instance, how does the novice "unlearn" incorrect knowledge? That is, if the novice's working model is incorrect, what happens when the inconsistencies are discovered? For that matter, how are the inconsistencies discovered in the first place?

CONCLUSIONS

By incorporating recent progress in causal reasoning into our working model, in addition to the traditional symptom-fault approach, we hope to produce a robust program for fault diagnosis in the automobile repair domain. Using protocols and other psychological methods while developing our model ensures that the computational diagnosis proceeds in a manner analogous to real mechanics. This will be invaluable later as the program is incorporated into a tutoring system.

ACKNOWLEDGEMENTS

Thanks to Janet Kolodner and Juliana Lancaster, primary investigators on this project, for their advice and guidance.

REFERENCES

Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman & J. S. Brown (Eds.) Intelligent Tutoring Systems. New York: Academic Press.

de Kleer, J. and Brown, J. S. (1981). Mental Models of Physical Mechanisms and their Acquisition. In J. R. Anderson (Ed.) Cognitive Skills and Their Acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Forbus, K. D. and Gentner, D. (1986). Learning Physical Domains: Toward A Theoretical Framework. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.) Machine Learning: An Artificial Intelligence Approach. Los Altos, CA: Morgan Kaufmann Publishers.

Hunt, R. M. (1981). Human Pattern Recognition and Information Seeking in Simulated Fault Diagnosis Tasks. Ph.D Thesis. Report T-110, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign.

Kolodner, J. L., Simpson, R. L., and Sycara-Cyranski, K. (1985). A Process Model of Case-Based Reasoning In Problem Solving. Proceedings of the International Joint Conference on Artificial Intelligence, Los Angeles, CA.

Kuipers, B. (1981). De Kleer and Brown's 'mental models': A critique. Working papers in cognitive science (17). Boston, MA: Tufts University.

Kuipers, B. (1984). Commonsense Reasoning About Causality: Deriving Behavior from Structure. Artificial Intelligence, 24, 169-203.

Kuipers, B. (1986). Qualitative Simulation as Causal Explanation. Technical Report AI TR86-24, Artificial Intelligence Laboratory, University of Texas at Austin.

Lehnert, W. G. (1978). Representing Physical Objects In Memory. Research Report #131. Department of Computer Science, Yale University, New Haven, CT.

Sembugamoorthy, V. and Chandrasekaran, B. (1986). Functional Representation of Devices and Compilation of Diagnostic Problem-Solving Systems. In J. L. Kolodner & C. K. Riesbeck (Eds.) Experience, Memory, and Reasoning. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shorliffe, E. H. (1976). Computer-based medical consultations: MYCIN. New York, NY: American Elsevier.

Simpson, R. L. (1985). A Computer Model of Case-Based Reasoning In Problem Solving: An Investigation in the Domain of Dispute Mediation. Ph. D. Thesis. Technical Report #GIT-ICS-85/18. School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA.

Sleeman, D. (1982). Assessing aspects of competence in basic algebra. In D. Sleeman & J. S. Brown (Eds.) Intelligent Tutoring Systems. New York: Academic Press.

Weld, D. (1986). The Use of Aggregation in Causal Simulation. Artificial Intelligence, 30, 1-34.